

SPEECH SPEED CHANGER

Publication number: JP2001350500 (A)

Publication date: 2001-12-21

Inventor(s): EBIHARA MITSURU; ISHIKAWA YASUSHI

Applicant(s): MITSUBISHI ELECTRIC CORP

Classification:

- international: G10L21/04; G10L13/06; G10L19/00; H03M7/30; G10L21/00; G10L13/00; G10L19/00; H03M7/30; (IPC1-7): G10L21/04; G10L13/06; G10L19/00; H03M7/30

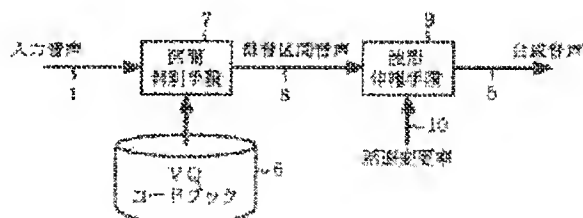
- European:

Application number: JP20000171005 20000607

Priority number(s): JP20000171005 20000607

Abstract of JP 2001350500 (A)

PROBLEM TO BE SOLVED: To obtain a speech speed changer which automatically conducts a high quality speech speed change by waveform expanding and compressing a vowel segment in accordance with the speech speed change rate specified by a user. SOLUTION: Code vectors relative to learning voice data are recorded in a VQ code book 6 with appearance probabilities of vowels and consonants of each code vector. Then, input voice 1 is vector quantized by a segment discriminating means 7 while referring to the VQ code book to select code vectors. Then, waveform compressing and expanding is performed for the segment which is discriminated to be the vowel segment of the selected code vector by employing a speech speed change rate 10 given to a waveform expanding and compressing means 9.



(11)特許出願公開番号
特開2001-350500
(P2001-350500A)

(51)Int.Cl. ⁷	識別記号	F I	テ-マ-ト*(参考)
G 1 0 L 21/04		H 0 3 M 7/30	B 5 D 0 4 5
13/06		G 1 0 L 3/02	A 5 J 0 6 4
19/00			A
H 0 3 M 7/30			E

BC14 BC28 BD01

```

graph LR
    1[入力音声 1] --> 7[区間判別手段 7]
    7 --> 8[区間音声 8]
    8 --> 9[波形圧縮手段 9]
    6[(VQコードブック 6)] --> 7
    10[ビットレート変更率 10] --> 9
    9 --> 5[合成音声 5]
  
```

【特許請求の範囲】

【請求項1】 学習音声データについてのコードベクトルを、前記各コードベクトルの母音または子音の出現確率とともに記録するベクトル量子化コードブックと、前記ベクトル量子化コードブックを参照し、入力音声についてベクトル量子化を行ってコードベクトルを選択し、選択された前記コードベクトルにおける母音または子音の出現確率に基づいて、当該区間が母音区間または子音区間のいずれであるかを判別する区間判別手段と、前記区間判別手段によって母音区間と判定された区間について、発話速度の変更の際に、元の音声に対する伸縮比率として与えられる話速変更率に応じて波形圧縮または伸長を行う波形伸縮手段とを備えた話速変更装置。

【請求項2】 波形伸縮手段が、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定し、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、母音区間と子音区間、あるいは母音区間と子音区間と無音区間の前記伸縮比に応じて、各区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行うものであることを特徴とする請求項1記載の話速変更装置。

【請求項3】 音素の種類や音素のカテゴリ毎に規定された固有の伸縮比を記憶する伸縮比テーブルを設け、波形伸縮手段が、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記伸縮比テーブルを参照して、入力音声の音素または音素のカテゴリ毎にその伸縮比に応じた区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行うものであることを特徴とする請求項1記載の話速変更装置。

【請求項4】 学習音声データについてのコードベクトルを、前記各コードベクトルに固有の伸縮比とともに記録する伸縮比付きベクトル量子化コードブックと、前記伸縮比付きベクトル量子化コードブックを参照し、入力音声についてベクトル量子化を行ってコードベクトルを選択し、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記選択したコードベクトル毎の伸縮比に応じた区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行うベクトル量子化波形伸縮手段とを備えた話速変更装置。

【請求項5】 入力音声の音素ラベルを抽出するラベリング手段と、前記入力音声の音素系列に固有のリズム知覚点を記憶するリズム知覚点テーブルと、前記リズム知覚点テーブルを参照して、前記音素ラベルから前記入力音声のリズム知覚点を抽出するリズム知覚点抽出手段と、

発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記リズム知覚点抽出手段で抽出されたリズム知覚点におけるリズム知覚点間隔の伸縮率が一定になるように区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行う波形伸縮手段とを備えた話速変更装置。

【請求項6】 波形伸縮手段が、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定し、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、リズム知覚点間隔の伸縮率が一定になるように、母音区間と子音区間、あるいは母音区間と子音区間と無音区間の前記伸縮比に応じて、各区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行うものであることを特徴とする請求項5記載の話速変更装置。

【請求項7】 音素の種類や音素のカテゴリ毎に規定された固有の伸縮比を記憶する伸縮比テーブルを設け、波形伸縮手段が、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記伸縮比テーブルを参照して、リズム知覚点間隔の伸縮率が一定になるように、入力音声の音素の種類または音素のカテゴリ毎に、その伸縮比に応じた区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行うものであることを特徴とする請求項5記載の話速変更装置。

【請求項8】 入力音声の音素ラベルを抽出するラベリング手段と、前記入力音声の音素系列に固有のリズム知覚点、および前記音素系列に固有の伸縮比を記憶するリズム知覚点兼伸縮比テーブルと、前記リズム知覚点兼伸縮比テーブルを参照して、前記音素ラベルから前記入力音声のリズム知覚点と前記音素系列毎の伸縮比とを抽出するリズム知覚点抽出手段と、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記リズム知覚点の間隔の伸縮率が一定になるように、前記音素系列毎の伸縮比に応じた各音素区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行う波形伸縮手段とを備えた話速変更装置。

【請求項9】 学習音声データについてのコードベクトルを、前記各コードベクトルの母音または子音の出現確率とともに記録するベクトル量子化コードブックと、前記ベクトル量子化コードブックを参照し、入力音声についてベクトル量子化を行ってコードベクトルを選択し、選択された前記コードベクトルにおける母音または子音の出現確率に基づいて、当該区間が母音区間または子音区間のいずれであるかを判別する区間判別手段と、前記母音区間の母音中心位置を決定する母音中心抽出手

段と、

前記母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定し、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記母音中心位置の間隔における伸縮率が一定になるように、母音区間と子音区間、あるいは母音区間と子音区間と無音区間の前記伸縮比に応じて、各区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行う波形伸縮手段とを備えた話速変更装置。

【請求項10】 学習音声データについてのコードベクトルを、前記各コードベクトルの母音または子音の出現確率とともに記録するベクトル量子化コードブックと、前記ベクトル量子化コードブックを参照し、入力音声についてベクトル量子化を行ってコードベクトルを選択し、選択された前記コードベクトルにおける母音または子音の出現確率に基づいて、当該区間が母音区間または子音区間のいずれであるかを判別する区間判別手段と、前記区間判別手段にて判定された母音区間の母音中心位置を決定する母音中心抽出手段と、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記母音区間内の波形伸縮率を、前記母音中心位置で最大または最小となり、子音区間との境界に近づくにつれ子音区間の波形伸縮率に近づくように決定して、その波形伸縮率に基づいて各区間の波形圧縮または伸長を行う母音中心部波形伸縮手段とを備えた話速変更装置。

【請求項11】 入力音声のピッチ周波数を分析し、ピッチ概形を得るピッチ分析手段と、前記ピッチ分析手段にて得られたピッチ概形の極大値と極小値を求めてピッチ概形基準位置を決定するピッチ概形基準位置決定手段とを設け、波形伸縮手段が、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定し、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記ピッチ概形基準位置の間隔の伸縮率が一定になるように、前記母音区間と子音区間の伸縮比、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行うものであることを特徴とする請求項1記載の話速変更装置。

【請求項12】 入力音声のピッチ周波数を分析し、ピッチ概形を得るピッチ分析手段と、区間判別手段にて子音区間と判定された暫定子音区間音声を、前記ピッチ分析手段によるピッチ概形の判定結果により、母音区間に含まれる部分と子音区間の部分とに分離し、分離された子音区間の部分のみを子音区間音声として出力し、母音区間に含まれる部分と前記区間判別手段にて母音区間と判定された暫定母音区間音声とを母音区間音声として出力するピッチ概形分割手段とを設

け、

波形伸縮手段が、ピッチ概形分割手段によって改めて母音区間および子音区間と判別された区間について固有の伸縮比を決定し、発話速度を変更するための、元の音声に対する伸縮比率による話速変更率が与えられた場合に、前記伸縮比に応じた区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形圧縮または伸長を行うものであることを特徴とする請求項1記載の話速変更装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、あらかじめ録音された音声、あるいはマイクなどから直接入力された音声の発話速度を、利用者の要求に応じて自由に変更することができる話速変更装置に関するものである。

【0002】

【従来の技術】近年、カーナビゲーションやPDA(Personal Digital Assistant)、電子玩具などの用途に適用可能な音声合成S/W(ソフトウェア)への要求が高まっている。こうしたアプリケーションでは、発話速度やピッチの変更などの多様性が、少ない演算量とメモリで実現されることが望まれている。こうした用途には小規模のテキスト音声合成S/Wの利用が考えられるが、実際のアプリケーションではテキスト音声合成のみが用いられることは稀で、合成音声の品質の観点から、録音された固定メッセージ、または固定メッセージと規則合成音声の併用がしばしば用いられている。

【0003】こうしたS/Wでは、固定メッセージは録音音声を蓄積し、これを再生するだけであるが、単語や句単位で録音された音声を接続したり、規則合成音声と接続して文としての連続性を得るためには、発話速度などの韻律に対する変更が必要となる。また、ユーザが自分の好みや聞き取りやすさに合わせて、発話速度を変更できることは有効であり、玩具などの応用としても要求は高い。

【0004】このように、原音声の発話速度をピッチや音調性を保ったまま変更する方式は既に検討されている。そのような従来の話速変更装置として、例えば特開平1-93795号公報に示された「音声の発話速度変換方法」などがある。図13は上記公報に示された手法に基づく、従来の話速変更装置の一構成例を示すブロック図である。図において、1は入力音声、2は分析部、3は制御部、4は波形接続部、5は合成音声である。

【0005】次に動作について説明する。入力音声1が分析部2に入力されると、分析部2ではその入力音声1について、有音と無音および有声音と無声音の判別が行われ、有声音については線形予測分析やピッチ分析が行われる。そして線形予測分析結果から共振周波数と帯域幅の時間変化を求め、それを利用して有声音を母音と子音に分離し、それを制御部3に入力する。制御部3で

は、子音は母音や無音より発話速度の変化が小さいという知識に基づいて、有声子音と無声子音の伸縮比率が母音や無音のそれより小さい値に設定される。波形接続部4ではこの制御部3で決定された伸縮比率によって各区間の発声時間長を伸縮して接続する。有声区間はピッチ周期を基にピッチ波形の繰り返しや間引きにより、また、無音区間長や無声区間は伸縮比率に応じた長さの波形の削除または繰り返し挿入によって発話速度を変更し、合成音声5として出力する。

【0006】なお、このような従来の話速変更装置に関連する記載のある文献としては、この他にも、例えば、合成音の母音長と子音長を決定する韻律生成規則に関するものとして、特開平6-266391号公報、特開平6-274195号公報、特開平10-78795号公報などがあり、母音区間と子音区間の波形振幅値を制御するものとして、特開平10-145897号公報が、別途指定する発話テンポに応じて合成音の母音長と子音長を決定する韻律生成規則に関するものとして、特開平6-222793号公報が、最初の母音長を検出して発話速度を決定するものとして、特開平10-70790号公報などがある。

【0007】

【発明が解決しようとする課題】従来の話速変更装置は以上のように構成されているので、母音と子音を区分して、それぞれを固有の波形伸縮率により波形圧縮または伸長を行う際に、線形予測分析結果を基に区分化しているため、母音と有声子音の判別に誤りが生じやすく、その境界も不明瞭となりがちになり、また、より自然な波形伸縮を行うためには子音の種類を考慮することが必要で、この枠組では不十分であり、さらに、分析部2において分析に要する演算量も多大で、実時間上での動作を考える場合には実現が難しいなどの課題があった。

【0008】この発明は上記のような課題を解決するためになされたもので、ベクトル量子化（Vector Quantization：以下、VQと略記することもある）されたパラメータ情報を用いて母音と子音を判別し、その判定結果を基に母音区間と子音区間毎に固有の波形伸縮率を与えて波形伸縮を行うことが可能であり、実時間で高品質な波形伸縮を実現することができる話速変更装置を得ることを目的とする。

【0009】

【課題を解決するための手段】この発明に係る話速変更装置は、VQコードブックに、学習音声データについてのコードベクトルを各コードベクトルの母音および子音の出現確率とともに記録しておき、区間判別手段において、そのVQコードブックを参照して入力音声をベクトル量子化してコードベクトルの選択を行い、選択されたコードベクトルにおける母音または子音の出現確率より、当該区間が母音区間か子音区間かの判別を行い、波形伸縮手段に話速変更率が与えられると、区間判別手段

にて母音区間と判定された区間について、与えられた話速変更率に基づいて波形伸縮を行うようにしたものである。

【0010】この発明に係る話速変更装置は、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比率を規定し、話速変更率が与えられると、規定された伸縮比に応じて、母音区間、子音区間、無音区間の各区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形伸縮を行う機能を波形伸縮手段に持たせたものである。

【0011】この発明に係る話速変更装置は、伸縮比テーブルを設けて、音素の種類やカテゴリ毎に規定された固有の伸縮比を記憶させ、話速変更率が与えられると、波形伸縮手段はその伸縮比テーブルを参照して、入力音声の音素毎または音素のカテゴリ毎の伸縮比に応じた区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うようにしたものである。

【0012】この発明に係る話速変更装置は、学習音声データのコードベクトルを各コードベクトルに固有の伸縮比率とともに記録しているVQコードブックを、VQ波形伸縮手段にて参照し、入力音声についてベクトル量子化してコードベクトルを選択し、話速変更率が与えられた場合に、その選択されたコードベクトル毎の伸縮比に応じた区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うようにしたものである。

【0013】この発明に係る話速変更装置は、入力音声の音素ラベルをラベリング手段で抽出し、リズム知覚点抽出手段において音素列毎に固有のリズム知覚点を記憶したリズム知覚点テーブルを参照して、音素ラベルから入力音声のリズム知覚点を抽出し、波形伸縮手段に話速変更率が与えられると、リズム知覚点間隔の伸縮率が一定になるように区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うようにしたものである。

【0014】この発明に係る話速変更装置は、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を波形伸縮手段によって規定し、その波形伸縮手段に話速変更率が与えられた場合に、リズム知覚点間隔の伸縮率が一定になるように、母音区間と子音区間の伸縮比、あるいは母音区間と子音区間と無音区間の伸縮比に応じた各区間毎の波形伸縮率を求め、それに基づいた波形伸縮を行うようにしたものである。

【0015】この発明に係る話速変更装置は、伸縮比テーブルを設けて、音素の種類や音素のカテゴリ毎に規定された固有の伸縮比を記憶させ、話速変更率が与えられると、波形伸縮手段はその伸縮比テーブルを参照して、リズム知覚点間隔の伸縮率が一定になるように、入力音声の音素毎または音素のカテゴリ毎の伸縮比に応じた区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うようにしたものである。

【0016】この発明に係る話速変更装置は、入力音声

の音素ラベルをラベリング手段で抽出し、リズム知覚点抽出手段にて、音素系列に固有のリズム知覚点と音素系列に固有の伸縮比とを記憶するリズム知覚点兼伸縮比テーブルを参照して、音素ラベルから入力音声のリズム知覚点と音素区間毎の伸縮比を抽出し、波形伸縮手段に話速変更率が与えられると、そのリズム知覚点間隔の伸縮率が一定になるように、音素区間毎の伸縮比に応じた各音素区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うようにしたものである。

【0017】この発明に係る話速変更装置は、区間判別手段によりVQコードブックを参照して母音区間音声抽出し、抽出された母音区間音声における母音中心位置を母音中心抽出手段において決定し、波形伸縮手段によって、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定して、話速変更率が与えられた場合に、その母音中心間隔の伸縮率が一定になるように、母音区間と子音区間の伸縮比、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うようにしたものである。

【0018】この発明に係る話速変更装置は、区間判別手段によりVQコードブックを参照して母音区間音声を抽出し、抽出された母音区間音声における母音中心位置を母音中心抽出手段において決定し、話速変更率が与えられた場合に、波形伸縮手段によって母音区間内の波形伸縮率を、母音中心位置では最大または最小、子音区間との境界に近くなるにつれ子音区間の波形伸縮率に近づくように決定し、それに基づいて各区間の波形伸縮を行うようにしたものである。

【0019】この発明に係る話速変更装置は、ピッチ分析手段による入力音声のピッチ周波数分析で得られたピッチ概形の極大値と極小値より、ピッチ概形基準位置をピッチ概形基準位置決定手段にて求め、話速変更率が与えられた場合に、波形伸縮手段によって、そのピッチ概形基準位置間隔の伸縮率が一定になるように、母音区間と子音区間の伸縮比、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うようにしたものである。

【0020】この発明に係る話速変更装置は、ピッチ分析手段のピッチ周波数分析結果によるピッチ概形の判定により、ピッチ概形分割手段で区間判別手段からの暫定子音区間音声を母音区間を含む部分と子音区間の部分とに分離して、改めて母音区間と子音区間との判別を行い、波形伸縮手段に話速変更率が与えられた場合に、その改めて母音区間と子音区間に判別された区間について固有の伸縮比を決定し、その伸縮比に応じて求めた区間毎の波形伸縮率に基づいて、波形伸縮を行うようにしたものである。

【0021】

【発明の実施の形態】以下、この発明の実施の一形態を

説明する。

実施の形態1. 図1はこの発明の実施の形態1による話速変更装置の一構成例を示すブロック図である。図において、1はこの話速変更装置に入力される、あらかじめシステムに用意された音声データによる入力音声であり、5はこの話速変更装置により発話速度の変更が行われて、外部に出力される合成音声である。6は学習音声データについてのコードベクトルを、各コードベクトルの母音や子音などの音素の出現確率とともに記録するVQコードブックである。7はこのVQコードブック6を参照し、入力された入力音声1についてベクトル量子化を行ってコードベクトルを選択し、その選択したコードベクトルにおける音素の出現確率に基づいて、当該区間が母音区間であるか子音区間であるかの判別を行う区間判別手段であり、8はこの区間判別手段7よりその判定結果に基づいて出力される母音区間音声である。9は区間判別手段7にて判定された母音区間音声8について、所望の話速変更率で波形圧縮または伸長を行い、合成音声5を生成して出力する波形伸縮手段であり、10はこの波形伸縮手段9に与えられる話速変更率である。

【0022】次に動作について説明する。ここで、入力音声1は音声を出力するシステムがあらかじめ用意した音声データであるものとする。また、VQコードブック6はこの入力音声1の音声データをベクトル量子化するために利用するものであり、学習音声データにおける音波形のスペクトルデータを一般的なクラスタリング手法によって分割したパターン（コードベクトル）と、そのコードの組とからなっている。なお、この実施の形態1におけるVQコードブック6においては、後で説明する母音の出現確率をコードベクトルとともに記憶している。入力音声1の入力を受けた区間判別手段7はこのVQコードブック6を参照して、入力音声1における母音と子音の区間を判別し、母音区間部分の入力音声1である母音区間音声8を波形伸縮手段9に送る。なお、このVQコードブック6を用いた区間判別の方法については後程説明する。

【0023】この波形伸縮手段9には上記母音区間音声8が入力されるとともに、話速変更率10も与えられる。この話速変更率10はユーザが音声の発話速度を変更する際に与える値で、元の音声に対する伸縮比率として波形伸縮手段9に入力される。例えば元の音声の β 倍（ $\beta > 1$ ）の速さで音声を出力したい場合には、この話速変更率 α は $\alpha = \beta$ となり、 β 倍の速さで出力したい場合には、話速変更率 α は $\alpha = 1/\beta$ となる。波形伸縮手段9はこの話速変更率10に応じて母音区間音声8の波形伸縮を行い、それ以外の区間についての波形伸縮は行わない。なお、この場合の母音区間の波形伸縮率は、母音区間長と子音区間長により次のように決められる。

【0024】例えば全音声区間L、母音区間長 L_v として、

$$\alpha = (\alpha' \cdot Lv + 1 \cdot (L - Lv)) / L$$

となるように、母音部分の波形伸縮率 α' を決定する。すなわち、この式を変形した次式より母音部分の波形伸縮率 α' が得られる。

$$\alpha' = 1 + (\alpha - 1) \cdot L / Lv$$

ちなみに、母音区間長 Lv は区間判別の時点で求められる。

【0025】ここで、波形伸縮手段9による母音区間音声8の伸縮方法としては、公知のPSOLA手法に基づく方法が一般的であるが、Oscillator Modelというピッチ依存性のない公知の方式を用いることも有効である。こうして波形伸縮手段9により話速変更された音声を合成音声5として出力する。

【0026】以下VQコードブック6を用いた区間判別について説明する。コードとそのコードベクトルのみからなるVQコードブック6を用いて、まず母音または子音としてラベル付けされた学習音声データのベクトル量子化を事前に行う。その時の量子化尺度はスペクトルパラメータである。次に、VQコードブック6の各コードベクトルに対して、母音または子音が選択された出現頻度を求め、この母音または子音の出現確率をコードベクトルとともにVQコードブック6に記録する。区間判別手段7はそのVQコードブック6を参照し、入力音声1についてスペクトル分析を行ってベクトル量子化する。このベクトル量子化によりVQコードブック6からコードベクトルを選択して、そこに記録されている出現確率から当該区間が母音区間または子音区間のいずれであるかを決定する。例えばある音声区間についてベクトル量子化により選択されたコードベクトルにおいて、そこに記録される母音の出現確率が70%であれば当該区間を母音区間と判別し、30%であれば子音区間と判別する。この例では母音区間と子音区間とを判別する閾値は50%としている。

【0027】以上のように、この実施の形態1によれば、母音区間についての波形伸縮を自動的に実施することが可能となって、全音声区間を一樣に同じ伸縮率で変更することによる自然性劣化の問題を解消することができ、良好な発話速度変更を行うことが可能になるという効果が得られる。

【0028】実施の形態2。図2はこの発明の実施の形態2による話速変更装置の一構成例を示すブロック図である。図において、1は入力音声、5は合成音声、6はVQコードブック、7は区間判別手段、8は母音区間音声、10は話速変更率であり、これらは図1に同一符号を付して示した、実施の形態1における各構成要素およびデータと同等の部分である。また、11は区間判別手段7よりその判定結果に基づいて出力される子音区間音声である。9は図1に同一符号を付して示した構成要素に相当する波形伸縮手段であるが、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸

縮比を規定し、話速変更率10が与えられた場合に、母音区間と子音区間の伸縮比、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率を求め、それに基づいた波形伸縮を行う点で、実施の形態1のそれとは異なっている。

【0029】次に動作について説明する。この場合も、入力音声1は音声を出力するシステムがあらかじめ用意した音声データであるものとする。また、VQコードブック6は音声データ1のベクトル量子化を行うために利用するものであり、その構成は実施の形態1の場合と同様である。区間判別手段7は入力音声1における母音と子音の区間を求め、母音区間部分の入力音声1を母音区間音声8として、子音区間部分の入力音声1を子音区間音声11としてそれぞれ波形伸縮手段9に送る。なお、上記区間判別例としては、実施の形態1の場合と同様の方法があげられる。

【0030】話速変更率10も実施の形態1の場合と同様に、ユーザが音声の発話速度を変更する際に与える値であり、元の音声に対する伸縮比率として波形伸縮手段9に入力される。なお波形伸縮手段9には母音区間と子音区間の伸縮比が規定されている。一般的に母音区間では子音区間より話速変化による伸縮の度合いが大きいことが知られている。伸縮比の値は発話速度の異なる学習音声データから得た母音全体および子音全体の平均区間長と発話速度との相関から求められる。

【0031】波形伸縮手段9にユーザより話速変更率10が与えられると、先に規定された母音区間と子音区間の伸縮比より、母音区間と子音区間毎の波形伸縮率が決定される。ここで、母音区間長を Lv 、子音区間長を Lc とすると、母音区間長 Lv と子音区間長 Lc のトータルの伸縮率は話速変更率 α ($\alpha > 0$) である。例えば母音区間の伸縮比を β 、子音区間の伸縮比を γ とすると、母音区間と子音区間の伸縮比については、波形伸縮の場合には $0 < \gamma < 1 < \beta$ 、波形圧縮の場合には $0 < \beta < 1 < \gamma$ であれば、母音区間の伸縮率が $\alpha' \beta$ 、子音区間の伸縮率が $\alpha' \gamma$ となるような係数 α' を求めることによって決まる。

【0032】すなわち、

$$\alpha = \alpha' (\beta \cdot Lv + \gamma \cdot Lc) / (Lv + Lc)$$

$$\alpha' = \alpha (Lv + Lc) / (\beta \cdot Lv + \gamma \cdot Lc)$$

より、母音区間の波形伸縮率は、

$$\alpha' \beta = \alpha \cdot \beta (Lv + Lc) / (\beta \cdot Lv + \gamma \cdot Lc)$$

子音区間の波形伸縮率は、

$$\alpha' \gamma = \alpha \cdot \gamma (Lv + Lc) / (\beta \cdot Lv + \gamma \cdot Lc)$$

となる。ちなみに、母音区間長 Lv と子音区間長 Lc は区間判別の際に求められるものとする。

【0033】このように、波形伸縮手段9は入力音声1の母音区間音声8と子音区間音声11のそれぞれについ

て、決まった波形伸縮率により伸縮を行う。その波形伸縮の方法は実施の形態1の場合と同様である。こうして波形伸縮手段9により話速変更された音声合成音声5として出力される。

【0034】以上のように、この実施の形態2によれば、母音区間と子音区間について、個別の伸縮率による波形伸縮を自動的に実施することができ、全音声区間を一樣に同じ伸縮率で変更することによる自然性劣化の問題を解消することが可能となって、良好な発話速度変更が実現できるようになるという効果が得られる。

【0035】実施の形態3。図3はこの発明の実施の形態3による話速変更装置の一構成例を示すブロック図である。なお、上記実施の形態1あるいは実施の形態2の各構成要素およびデータに相当する部分については、図1または図2と同一符号を付してその説明を省略する。図において、12は音素の種類や、音素カテゴリー毎に規定された固有の伸縮比を記憶する伸縮比テーブルである。なお、波形伸縮手段9は図1または図2に同一符号を付して示した実施の形態1あるいは実施の形態2の構成要素に相当する波形伸縮手段であるが、話速変更率10が与えられた場合に、伸縮比テーブル12を参照して、入力音声1の音素毎または音素カテゴリー毎にその伸縮比に応じた区間毎の波形伸縮率を求めて波形伸縮を行う点でそれとは異なっている。

【0036】次に動作について説明する。この場合も、入力音声1は音声出力するシステムがあらかじめ用意した音声データ1であるものとする。VQコードブック6は音声データのベクトル量子化を行うために利用するものであり、その構成は実施の形態1の場合と同様である。区間判別手段7は入力音声1における母音と子音の区間を求め、母音区間部分の入力音声1である母音区間音声8と、子音区間部分の入力音声1である子音区間音声11を波形伸縮手段9に送る。ただし、子音区間音声は後に説明する音素カテゴリー毎に分類されている。なお、上記区間判別の例も実施の形態2の場合と同様である。

【0037】話速変更率10も実施の形態1の場合と同様に、ユーザが音声の発話速度を変更する際に与える値であり、元の音声に対する伸縮比率として波形伸縮手段9に入力される。なお伸縮比テーブル12には前述のように、音素カテゴリー毎の伸縮比が記述されている。この音素カテゴリーの例としては、母音、摩擦性子音、破裂性子音、鼻音、流音、撥音などが挙げられ、伸縮比にばらつきがある子音区間をこのように分類して、音素カテゴリー毎にその伸縮比を決定する。このカテゴリー毎の伸縮比の値は、発話速度の異なる学習音声データから得た、音素カテゴリー毎の平均時間長と発話速度との関係から求められる。

【0038】波形伸縮手段9にユーザより話速変更率10が与えられると伸縮比テーブル12を参照し、当該伸

縮比テーブル12に記憶されている音素の種類や音素カテゴリー毎に固有の伸縮比に基づいて、母音区間と子音区間の波形伸縮率を決定する。なお、この母音区間と子音区間に与えられた伸縮比からの波形伸縮率の決定法は、実施の形態2の場合と同様である。このように、波形伸縮手段9は母音区間と子音カテゴリー毎に分類された区間それぞれについて、入力音声1の伸縮を行う。この波形伸縮の方法は実施の形態1の場合と同様である。こうして波形伸縮手段9により話速変更された音声合成音声5として出力する。

【0039】以上のように、この実施の形態3によれば、母音区間と子音カテゴリー毎に分類された各区間について、個別の伸縮率による波形伸縮を自動的に実施することが可能になって、全音声区間を一樣に同じ伸縮率で変更することによる自然性劣化の問題を解消することができ、良好な発話速度変更の実現が可能になるという効果が得られる。

【0040】実施の形態4。図4はこの発明の実施の形態4による話速変更装置の一構成例を示すブロック図である。図において、1は入力音声、5は合成音声、10は話速変更率であり、これらは図1から図3に同一符号を付して示した、実施の形態1～実施の形態3における各構成要素およびデータと同等のものである。13は学習音声データについてのコードベクトルを、各コードベクトルの固有の伸縮比とともに記録する伸縮比付きVQコードブックである。14は入力音声1について、この伸縮比付きVQコードブック13を参照してベクトル量子化を行って、ベクトルコードを選択し、話速変更率10が与えられた場合に、選択したコードベクトル毎の伸縮比に応じた区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行うVQ波形伸縮手段である。

【0041】次に動作について説明する。この場合も、入力音声1は音声出力するシステムがあらかじめ用意した音声データであるものとする。また、伸縮比付きVQコードブック13はその音声データのベクトル量子化を行うために利用するものであり、その構成は後に説明する。話速変更率10も実施の形態1の場合と同様であり、ユーザが音声の発話速度を変更する際に与える値で、元の音声に対する伸縮比率としてVQ波形伸縮手段14に入力される。VQ波形伸縮手段14は伸縮比付きVQコードブック13を参照して、そこに記憶されている伸縮比と、ユーザから与えられる話速変更率10とによって波形伸縮を行う。なお、このVQ波形伸縮手段14における波形伸縮法は後に説明する。このようにして、VQ波形伸縮手段14により話速変更された音声合成音声5として出力される。

【0042】以下伸縮比付きVQコードブック13の構成を説明する。この音声データのベクトル量子化に利用される伸縮比付きVQコードブック13は、学習音声データにおける音声波形のスペクトルデータを、一般的に

クラスタリング手法によって分割したパターン（コードベクトル）と、そのコードの紐からなっている。この伸縮比付きVQコードブック13については事前に、母音または子音としてラベル付けされた学習音声データのベクトル量子化を行っておく。その時の量子化尺度はスペクトルパラメータである。

【0043】次に、伸縮比付きVQコードブック13の各コードベクトルに対して、母音または子音が選択された出現頻度を求め、母音区間または子音区間の出現確率をコードベクトルとともに伸縮比付きVQコードブック13に記録する。さらに、各コードベクトル毎に記録された母音区間および子音区間の出現確率から、各コードベクトルが母音に対応するものか子音に対応するものかを判定しておき、それぞれのコードベクトル毎に母音区間または子音区間に固有の伸縮比を与えて記憶する。例えば、母音区間と子音区間とを判別する閾値を50%として、コードベクトルの母音出現確率が70%であれば母音区間の波形伸縮率を、30%であれば子音区間の波形伸縮率をそれぞれ記録する。ここでの伸縮比は実施の形態2のものと同様とする。

【0044】次に、VQ波形伸縮手段14における波形伸縮について説明する。VQ波形伸縮手段14はまず、伸縮比付きVQコードブック13を参照し、入力音声1についてスペクトル分析を行ってそれをベクトル量子化する。このベクトル量子化に基づいて伸縮比付きVQコードブック13からコードベクトルを選択し、そこに記憶される伸縮比を読み出す。ユーザより話速変更率10が与えられると、伸縮比付きVQコードブック13から読み出した伸縮比により、母音区間と子音区間毎の波形伸縮率を決定する。なお、この波形伸縮率の決定方法は実施の形態2の場合と同様である。これらの波形伸縮率に応じて当該区間の入力音声1の波形伸縮を行い、それを合成音声5として出力する。なお、この波形伸縮の方法も実施の形態1の場合と同様である。

【0045】以上のように、この実施の形態4によれば、コードベクトルに固有の伸縮比率を記録することによって、母音区間と子音区間についての個別の伸縮率での波形伸縮を簡便かつ自動的に実施することができ、全音声区間を一律に同じ伸縮率で変更することによる自然性劣化の問題を解消することが可能となって、良好な発話速度変更を実現することができるという効果が得られる。

【0046】実施の形態5。図5はこの発明の実施の形態5による話速変更装置の一構成例示ブロック図である。なお、上記実施の形態1～実施の形態4の各構成要素およびデータに相当する部分については、図1～図4と同一符号を付してその説明を省略する。図において、15は入力音声1における音素境界と音素種類を判別し、その識別結果より音素ラベルを抽出するラベリング手段であり、16はこのラベリング手段15によって抽

出された上記音素ラベルである。17は音素系列に固有のリズム知覚点を記憶するリズム知覚点テーブルであり、18はこのリズム知覚点テーブル17を参照して、ラベリング手段15の抽出した音素ラベル16から、入力音声1のリズム知覚点を抽出するリズム知覚点抽出手段である。なお、波形伸縮手段9は図1～図3に同一符号を付して示したも実施の形態1～実施の形態3の構成要素に相当するものであるが、話速変更率10が与えられた場合に、リズム知覚点間隔の伸縮率が一定になるように区間毎の波形伸縮率を求め、それに基づいて波形伸縮を行っている点で、上記各実施の形態における波形伸縮手段9とは異なっている。

【0047】次に動作について説明する。この場合も、入力音声1は音声を出力するシステムがあらかじめ用意した音声データであるものとする。入力音声1が入力されると、ラベリング手段15はその入力音声1における音素境界と音素種類の判別を行い、その判別結果に基づいて音素ラベル16を抽出し、それをリズム知覚点抽出手段18へ送る。なお、この音素ラベル16の決定方法としては、例えば、入力音声1とその発話内容から、事前にマニュアルで音素境界ならびに音素種類を与えておく方法がある。

【0048】一方、リズム知覚点テーブル17には子音（C）や母音（V）の組合せに応じたリズム知覚点の情報が記憶されている。リズム知覚点は人間が聴覚的にリズムを知覚する音声波形上の位置であり、CV（子音ー母音）またはVCV（母音ー子音ー母音）などの音素系列毎に固有の値を取る。このリズム知覚点の間隔が自然音声のリズムと対応しており、発話速度が変わっても同じ音素の組合せであれば、そのリズム知覚点間隔間の比率は保存されている。

【0049】このリズム知覚点テーブル17の内容の一例を図6に示す。リズム知覚点テーブル17には図6

(a)に示すように、音素系列とその音素系列に対応するリズム知覚点が記述されている。すなわち、図6

(a)の例では先行V、C、後続VによるCVもしくはVCVの音素系列に対応するリズム知覚点が記述されている。また、この図6(a)の例では、リズム知覚点はCVまたはVCVの音声素片におけるフレーム位置を示しており、音声素片のパワー概形との対応においては図6(b)に示す位置となる。リズム知覚点抽出手段18は音素ラベル16の組からこのリズム知覚点テーブル17を参照して、入力音声1上のリズム知覚点位置を決定する。

【0050】話速変更率10も実施の形態1の場合と同様に、ユーザが音声の発話速度を変更する際に与える値であり、元の音声に対する伸縮比率として波形伸縮手段9に入力される。波形伸縮手段9はこの話速変更率10に応じて入力音声1を伸縮するが、その際、入力音声1上のリズム知覚点間隔が同じ話速変更率10によって伸

縮されるようにする。すなわち、まずラベリング手段15にて抽出された音素ラベル16によって、リズム知覚点間隔内の母音区間と子音区間の内訳を判定する。なお、波形伸縮手段9で規定されている母音と子音の伸縮比から、母音区間と子音区間の波形伸縮率を決定する方法は実施の形態2の場合と同様である。波形伸縮手段9はこれらの値に応じて各区間の入力音声1を圧縮または伸長する。この波形伸縮の方法も実施の形態1の場合と同様である。こうして波形伸縮手段9により話速変更された音声は合成音声5として出力される。

【0051】以上のように、この実施の形態5によれば、波形伸縮を自動的に実施することが可能となり、自然音声の時間構造の基本であるリズム知覚点を考慮した波形伸縮を行うことによって、全音声区間を一様に変換するよりも自然な発話速度変更を実現することができるという効果が得られる。

【0052】実施の形態6、図7はこの発明の実施の形態6による話速変更装置の一構成例を示すブロック図である。なお、上記実施の形態1～実施の形態5の各構成要素およびデータに相当する部分については、図1～図5と同一符号を付してその説明を省略する。この実施の形態6による話速変更装置は実施の形態5による話速変更装置の波形伸縮手段9に伸縮比テーブル12を接続したものであり、波形伸縮手段9に話速変更率10が与えられた場合に、伸縮比テーブル12を参照して波形伸縮を行う点で、実施の形態5とは異なっている。

【0053】次に動作について説明する。この場合も、入力音声1は音声を出力するシステムがあらかじめ用意した音声データであるものとする。ラベリング手段15は入力音声1における音素境界と音素種類を判別し、その結果を音素ラベル16としてリズム知覚点抽出手段18に送る。なお、音素ラベル16の決定は実施の形態5の場合と同様とする。リズム知覚点テーブル17は音素（CやV）の組合せに応じたリズム知覚点の情報が記憶されているテーブルであり、その内容および構成は図6に示した実施の形態5の場合と同様である。リズム知覚点抽出手段18はラベリング手段15の抽出した音素ラベル16の組からリズム知覚点テーブル17を参照し、入力音声1上のリズム知覚点位置を決定する。

【0054】話速変更率10も実施の形態1の場合と同様に、ユーザが音声の発話速度を変更する際に与える値であり、元の音声に対する伸縮比率として波形伸縮手段9に入力される。波形伸縮手段9はこの話速変更率10に応じて入力音声1を伸縮するが、その際、入力音声1上のリズム知覚点間隔が同じ話速変更率10によって伸縮されるようにする。すなわち、まずラベリング手段15の抽出した音素ラベル16により、リズム知覚点間隔内の母音や破裂性子音などの音素カテゴリー毎の区間の内訳が判定される。この音素カテゴリーの例は実施の形態3において説明したものと同様である。そして伸縮比

テーブル12を参照することにより、区間毎の音素カテゴリーに対応する伸縮比を求める。この伸縮比テーブル12の構成と波形伸縮率の決定方法も、実施の形態3の場合と同様である。波形伸縮手段9はこれらの値に基づいて、各区間の入力音声1を圧縮もしくは伸長する。なお、この波形伸縮の方法も実施の形態1の場合と同様である。こうして波形伸縮手段9により話速変更された音声は合成音声5として出力される。

【0055】以上のように、この実施の形態6によれば、波形伸縮を自動的に実施することが可能となり、自然音声の時間構造の基本であるリズム知覚点と音素種類を考慮した波形伸縮を行うことによって、全音声区間を一様に変換するよりも自然な発話速度変更を実現することができるという効果が得られる。

【0056】実施の形態7、図8はこの発明の実施の形態7による話速変更装置の一構成例を示すブロック図であり、上記実施の形態1～実施の形態6の各構成要素およびデータに相当する部分については、図1～図5および図7と同一符号を付してその説明を省略する。図において、19は音素系列に固有のリズム知覚点と音素系列に固有の伸縮比とを記憶し、リズム知覚点抽出手段18によって参照されるリズム知覚点兼伸縮比テーブルである。上記各実施の形態において個別に用意されていた伸縮比テーブル12とリズム知覚点テーブル17が、この実施の形態7では1つのリズム知覚点兼伸縮比テーブル19にまとめられている点で、実施の形態6とは異なっている。

【0057】次に動作について説明する。この場合も、入力音声1は音声を出力するシステムがあらかじめ用意した音声データであるものとする。ラベリング手段15は入力音声1における音素境界と音素種類の判別を行い、その判別結果を音素ラベル16としてリズム知覚点抽出手段18へ送る。なお、この音素ラベル16の決定は実施の形態5の場合と同様とする。また、リズム知覚点兼伸縮比テーブル19には音素（CやV）の組合せに応じたリズム知覚点の情報と、その組合せに対応する音素波形の伸縮比が記憶されている。このリズム知覚点の内容は、図6に示した実施の形態5の場合と同様である。波形の伸縮比率はCVまたはVCVの各音素区間に対して伸縮比率を求め、それを記録したものである。

【0058】リズム知覚点抽出手段18は、ラベリング手段15にて抽出された音素ラベル16の組に基づいてリズム知覚点兼伸縮比テーブル19を参照し、入力音声1上のリズム知覚点位置を決定して音素系列に対応した伸縮比を得る。なお、話速変更率10は実施の形態1の場合と同様であり、ユーザが音声の発話速度を変更する際に与える値で、元の音声に対する伸縮比率として波形伸縮手段9に入力される。波形伸縮手段9はその話速変更率10に応じて入力音声1を伸縮するが、その際に、入力音声1上のリズム知覚点間隔が同じ話速変更率10

で伸縮されるようにする。

【0059】すなわち、リズム知覚点間隔内の母音や破裂性子音などの音素カテゴリー毎に区間の内訳を、音素ラベル16によりまず判定する。この音素カテゴリーの例は実施の形態3の場合と同様である。このようにしてリズム知覚点兼伸縮比テーブル19を参照することによって得た伸縮比と、与えられた話速変更率10から各区間毎の波形伸縮率を求める。なお、波形伸縮率の決定方法は実施の形態3の場合と同様である。波形伸縮手段9はこれらの値に応じて、各区間の入力音声1を伸縮する。波形伸縮の方法は実施の形態1の場合と同様である。こうして波形伸縮手段9により話速変更された音声は合成音声5として出力される。

【0060】以上のように、この実施の形態7によれば、波形伸縮を自動的に実施することが可能となり、自然音声の時間構造の基本であるリズム知覚点と音素種類を考慮した波形伸縮を行うことによって、全音声区間を一樣に変換するよりも自然な発話速度変更を実現することができるという効果が得られる。

【0061】実施の形態8、図9はこの発明の実施の形態8による話速変更装置の一構成例を示すブロック図である。図において、1は入力音声、5は合成音声、6はVQコードブック、7は区間判別手段、8は母音区間音声、9は波形伸縮手段、10は話速変更率であり、図1～図5および図7、図8に同一符号を付して示した、実施の形態1～実施の形態7における各構成要素およびデータと同等の部分である。

【0062】また、20は区間判別手段7が判別した母音区間の中心位置を決定する母音中心抽出手段であり、21はその母音中心抽出手段20によって抽出された母音中心位置である。なお、波形伸縮手段9は母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定し、話速変更率10が与えられた場合に、母音中心抽出手段20の抽出した母音中心位置21について、母音中心間隔の伸縮率が一定になるように、母音区間と子音区間の伸縮比、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率を求め、その波形伸縮率に基づいて波形伸縮を行っている点で、図1などに同一符号を付して示した上記各実施の形態における波形伸縮手段とは異なっている。

【0063】次に動作について説明する。この場合も、入力音声1は音声出力システムがあらかじめ用意した音声データであるものとする。VQコードブック6は音声データのベクトル量子化を行うために利用するものであり、その構成は実施の形態1におけるそれと同様である。区間判別手段7はこのVQコードブック6を参照して入力音声1における母音と子音の区間を求め、母音区間における入力音声1である母音区間音声8を母音中心抽出手段20に送る。なお、この区間判別の例としては、実施の形態1の場合と同様の方法があげられる。

母音区間音声8を受けた母音中心抽出手段20は、その母音区間内の基準点である母音中心位置21を抽出し、波形伸縮手段9に送出する。この母音中心位置抽出方法の例としては、母音区間内でパワーが最大となる位置を決定する方法などがある。

【0064】話速変更率10も実施の形態1の場合と同様であり、ユーザが音声の発話速度を変更する際に与える値で、元の音声に対する伸縮比率として波形伸縮手段9に入力される。波形伸縮手段9はその話速変更率10に応じて母音区間音声8および子音区間音声11を伸縮するが、その際に、母音区間音声8上の母音中心位置21の中心位置間隔が同じ話速変更率10で伸縮されるようにする。なお、波形伸縮手段9で規定されている母音と子音の伸縮比から母音区間と子音区間の波形伸縮率を決定する方法は、実施の形態2の場合と同様である。波形伸縮手段9はこれらの値に応じて、各区間の入力音声1を圧縮または伸長する。この波形伸縮の方法も実施の形態1の場合と同様である。こうして波形伸縮手段9により変更された音声は合成音声5として出力される。

【0065】以上のように、この実施の形態8によれば、波形伸縮を自動的に実施することが可能となり、自然音声の母音間の間隔を利用できるように母音中心位置を考慮した波形伸縮を行うことによって、全音声区間を一樣に変換するよりも自然な発話速度変更を実現することができるという効果が得られる。

【0066】実施の形態9、図10はこの発明の実施の形態9による話速変更装置の一構成例を示すブロック図であり、上記実施の形態1～実施の形態8の各構成要素およびデータに相当する部分については、図1～図5および図7～図9と同一符号を付してその説明を省略する。図において、22は話速変更率10が与えられた場合に、母音区間内の波形伸縮率を、母音中心抽出手段20が抽出した母音中心位置21で最大または最小となり、子音区間との境界に近くなるにつれてその子音区間の波形伸縮率に近付くように決定し、その決定された波形伸縮率に基づいて各区間の波形の伸縮を行う母音中心部波形伸縮手段である。実施の形態9は波形伸縮手段9をこの母音中心部波形伸縮手段22で代替している点で、実施の形態8とは異なっている。

【0067】次に動作について説明する。この場合も、入力音声1は音声出力システムがあらかじめ用意した音声データであるものとする。また、VQコードブック6も音声データのベクトル量子化を行うために利用するものであり、その構成は実施の形態1の場合と同様である。区間判別手段7はこのVQコードブック6を参照して入力音声1における母音と子音の区間を求め、母音区間における入力音声1である母音区間音声8を母音中心抽出手段20に送る。なお、区間判別の例としては、実施の形態1の場合と同様の方法があげられる。母音区間音声8を受けた母音中心抽出手段20は、その母

音区間内の基準点である母音中心位置21を抽出し、母音中心部波形伸縮手段22に送出する。この母音中心位置21とその抽出方法は、実施の形態8の場合と同様とする。

【0068】話速変更率10も実施の形態1の場合と同様であり、ユーザが音声の発話速度を変更する際に与える値で、元の音声に対する伸縮比率として母音中心部波形伸縮手段22に入力される。母音中心部波形伸縮手段22はこの話速変更率10に応じて母音区間音声8および子音区間音声11を伸縮するが、その際に入力音声1上の母音中心間隔が同じ話速変更率10で伸縮されるようにする。なお、母音区間と子音区間のそれぞれの波形伸縮率は、実施の形態8の場合と同様に決定することができる。

【0069】ここで、母音区間については子音区間との境界付近などの過渡部分と、母音中心位置21などの定常部分を同じ伸縮率で伸縮することは、音声の自然性を損なう可能性がある。そこで母音区間の波形伸縮率を、母音中心抽出手段20の抽出した母音中心位置21において最大または最小となり、子音区間との境界に近くなるにつれてその子音区間の波形伸縮率に近づくように決定する。すなわち、母音区間の伸縮率が $\lambda > 1$ である場合は波形伸長であるので、母音中心位置21で波形伸縮率を最大とし、子音区間との境界部分では λ より小さい当該子音区間の波形伸縮率と等しい値にまで減少させる。また母音区間の伸縮率が $0 < \lambda < 1$ の場合は波形伸縮であるので、母音中心位置21で波形伸縮率を最小とし、子音区間との境界部分では λ より大きい当該子音区間の波形伸縮率まで増加させる。母音中心部波形伸縮手段22はこの波形伸縮率に応じて、各区間の入力音声1を伸縮する。なお、波形伸縮の方法は実施の形態1の場合と同様である。こうして母音中心部波形伸縮手段22により変更された音声は合成音声5として出力される。

【0070】以上のように、この実施の形態9によれば、波形伸縮を自動的に実施することが可能となり、自然音声の母音間の間隔を利用できるように母音中心位置を考慮し、さらに子音との境界付近より母音中心位置周辺を強く伸縮するように重み付けをして波形伸縮を行うことによって、全音声区間を一様に変換するよりも自然な発話速度変更を実現することができるという効果が得られる。

【0071】実施の形態10。図11はこの発明の実施の形態10による話速変更装置の一構成例を示すブロック図である。なお、上記実施の形態1～実施の形態9における各構成要素およびデータに相当する部分については、図1～図5、図7～図10と同一符号を付してその説明を省略する。図において、23は入力音声1のピッチ周波数を分析し、ピッチ波形を得るピッチ分析手段であり、24はこのピッチ分析手段23によって抽出され

たピッチ波形である。25はそのピッチ波形24の最大値と最小値を求め、それに基づいてピッチ波形基準位置を決定するピッチ波形基準位置決定手段であり、26はこのピッチ波形基準位置決定手段25において決定されたピッチ波形基準位置である。

【0072】なお、波形伸縮手段9は、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定し、話速変更率10が与えられた場合に、ピッチ波形基準位置間隔の伸縮率が一定になるように、母音区間と子音区間の伸縮比、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率を求め、その波形伸縮率に基づいて区間判別手段7からの母音区間音声8および子音区間音声11の波形伸縮を行っているものである点で、図1～図5および図7～図10に同一符号を付して示した、上記各実施の形態における波形伸縮手段9とは異なっている。

【0073】次に動作について説明する。この場合も、入力音声1は音声を出力するシステムがあらかじめ用意した音声データであるものとする。ピッチ分析手段23は入力音声1が入力されると、その短時間のピッチ周波数を求め、それを時間方向に記録したピッチ波形24を抽出して、ピッチ波形基準位置決定手段25に入力する。なお、このピッチ周波数の抽出方法としては、一定区間長毎に音声波形からケプストラムを求め、それが高ケプレンシ領域で最大となる値を得、その逆数より求める方法がある。ピッチ波形基準位置決定手段25は、入力されたピッチ波形24上の基準位置であるピッチ波形基準位置26を決定し、それを波形伸縮手段9に出力する。ここで、このピッチ波形基準位置26は、例えば、大域的にピッチ波形24が最大値および最小値を取る時間位置とする。

【0074】また、VQコードブック6も音声データのベクトル量子化を行うために利用するものであり、その構成は実施の形態1の場合と同様である。区間判別手段7は入力音声1における母音と子音の区間を求め、その母音区間部分の入力音声1である母音区間音声8と、子音区間部分の入力音声1である子音区間音声11を波形伸縮手段9に送る。なお、この区間判別の例としては、実施の形態1の場合と同様の方法があげられる。

【0075】話速変更率10は実施の形態1の場合と同様であり、ユーザが音声の発話速度を変更する際に与える値で、元の音声に対する伸縮比率として波形伸縮手段9に入力される。波形伸縮手段9はこの話速変更率10に応じて入力音声1を伸縮するが、その際に、ピッチ波形基準位置決定手段25にて決定された入力音声1上のピッチ波形基準位置26の間隔が同じ話速変更率10で伸縮されるように線形伸縮する。すなわち、ユーザより話速変更率10が与えられると波形伸縮手段9はまず、母音区間と子音区間毎の波形伸縮率を決定する。なお、この波形伸縮率の決定方法は実施の形態2の場合と同様

である。波形伸縮手段9は次に、決定された波形伸縮率で母音区間と子音区間それぞれについて入力音声1を伸縮する。この波形伸縮の方法も実施の形態1の場合と同様である。こうして波形伸縮手段9により話速変更された音声は合成音声5として出力される。

【0076】以上のように、この実施の形態10によれば、波形伸縮を自動的に実施することが可能となり、自然音声のピッチ周波数より求めた基準位置に基づいて波形伸縮を行うことにより、ピッチ概形を破壊せずに波形伸縮を行うことができ、全音声区間を一樣に変換するよりも自然な発話速度変更の実現が可能になるという効果が得られる。

【0077】実施の形態11、図12はこの発明の実施の形態11による話速変更装置の一構成例を示すブロック図である。なお、上記実施の形態1～実施の形態10における各構成要素およびデータに相当する部分については、図1～図5および図7～図11と同一符号を付してその説明を省略する。図において、27は区間判別手段7が入力音声1より求めた母音区間における暫定母音区間音声であり、28は区間判別手段7が入力音声1より求めた子音区間における暫定子音区間音声である。29は区間判別手段7で判定された暫定子音区間音声28を、ピッチ分析手段23によるピッチ概形24の判定結果に基づいて、母音区間に含まれる部分と子音区間の部分とに分離し、その子音区間の部分を子音区間音声11として、母音区間に含まれる部分と暫定母音区間音声27とを母音区間音声8として波形伸縮手段9に出力するピッチ概形分割手段である。

【0078】なお、波形伸縮手段9は、ピッチ概形分割手段29によって改めて判別された母音区間音声8と子音区間音声11について、それぞれ固有の伸縮比を決定し、話速変更率10が与えられた場合に、この伸縮比に応じた区間毎の波形伸縮率を求めて、それに基づく波形伸縮を行うものである点で、図1～図5および図7～図11に同一符号を付して示した、上記各実施の形態における波形伸縮手段9とは異なっている。

【0079】次に動作について説明する。この場合も、入力音声1は音声を入力するシステムがあらかじめ用意した音声データであるものとする。また、VQコードブック6も音声データのベクトル量子化を行うために利用するものであり、その構成は実施の形態1の場合と同様である。区間判別手段7は入力音声1における母音と子音の区間を求め、母音区間の入力音声1を暫定母音区間音声27として、子音区間の入力音声1を暫定子音区間音声28として、それぞれピッチ概形分割手段29に出力する。なお、この区間判別の例としては、実施の形態1と同様の方法があげられる。

【0080】また、ピッチ分析手段23は入力音声1の短時間のピッチ周波数を求めて、それを時間方向に記録したものをピッチ概形24として抽出し、それをピッチ

概形分割手段29に送る。なお、このピッチ分析の例としては、実施の形態10と同様の方法があげられる。

【0081】区間判別手段7から暫定母音区間音声27と暫定子音区間音声28とが入力され、ピッチ分析手段23からピッチ概形24が入力されたピッチ概形分割手段29は、暫定子音区間音声28におけるピッチ概形24の分割を行う。すなわち、暫定子音区間音声28のピッチ概形24がランダムであったり、前後の暫定母音区間音声27のピッチ概形24と比べて概形の傾きが異なる場合には、その区間を子音区間音声11とし、暫定子音区間音声28のそれ以外の区間を暫定母音区間音声27とともに母音区間音声8とする。波形伸縮手段9には、このようにして、ピッチ概形分割手段29により改めて判別された母音区間音声8および子音区間音声11が入力される。

【0082】波形伸縮手段9は話速変更率10が与えられると、それに基づいて、ピッチ概形分割手段29からの母音区間音声8と子音区間音声11の固有の伸縮比を決定し、その伸縮比に応じた区間毎の波形伸縮率を求めて、それに基づいた波形の圧縮あるいは伸長を行う。なお、話速変更率10は実施の形態1の場合と同様であり、ユーザが音声の発話速度を変更する際に与える値で、元の音声に対する伸縮比率として波形伸縮手段9に入力される。また、波形伸縮率の決定方法も実施の形態2の場合と同様であり、さらに、波形伸縮の方法も実施の形態1の場合と同様である。こうして波形伸縮手段9により変更された音声は合成音声5として出力される。

【0083】以上のように、この実施の形態11によれば、波形伸縮を自動的に実施することが可能となり、ピッチ概形を母音子音区間判別の判定基準として利用することによって、ピッチ概形を破壊せずに波形伸縮を行うことができ、全音声区間を一樣に変換するよりも自然な発話速度変更の実現が可能になるという効果が得られる。

【0084】なお、実施の形態1～実施の形態11における入力音声1は、あらかじめシステムに用意された音声データである代りに、マイクによりナレータが直接入力した音声であってもよい。

【0085】また、実施の形態2～実施の形態11における母音区間の波形伸縮率は、ユーザが入力した話速変更率10をそのまま用いてもよく、子音区間の波形伸縮率を“1”に固定してもよい。

【0086】また、実施の形態1～実施の形態3、および実施の形態8～実施の形態11における区間判別手段7は、母音区間と子音区間の判別の他に、入力音声1の波形の振幅の絶対値が任意の値をすべて下回る区間を無音区間と判別する機能を備えるようにしてもよく、また、実施の形態1～実施の形態3、実施の形態8、実施の形態10、および実施の形態11の波形伸縮手段9と、実施の形態9の母音中心部波形伸縮手段22は、無

音区間を伸縮することができ、伸縮する際には母音区間と子音区間より無音区間の伸縮比率を最も大きく設定することができるようにしてもよい。

【0087】また、実施の形態1～実施の形態4、および実施の形態8～実施の形態11における区間判別手段7は、その都度スペクトルパラメータを分析する代りに、事前に分析したスペクトルパラメータをデータとして保持しておくようにしてもよい。

【0088】また、実施の形態1～実施の形態3、および実施の形態8～実施の形態11における区間判別手段7は、単一の閾値により母音と子音を一意に決定する代りに、例えば50%の付近の母音出現確率で、前後の区間が母音であれば母音と判別するなど、前後の区間との連続性を考慮した区間判別を行うようにしてもよい。

【0089】また、実施の形態1～実施の形態3、および実施の形態8～実施の形態11における区間判別手段7は、個々の区間で母音と子音を一意に決定する代りに、前後数区間の確率の累積値により判定するようにしてもよい。

【0090】また、実施の形態1～実施の形態4は、入力音声1そのものの代りに、入力音声を線形予測分析して得た残差波形とスペクトルパラメータを入力音声1として入力するようにしてもよく、また、実施の形態1および実施の形態3における区間判別手段7と、実施の形態4におけるVQ波形伸縮手段14は、ベクトル量子化と区間判別をスペクトルパラメータについて行い、実施の形態1～実施の形態3における波形伸縮手段9と実施の形態4におけるVQ波形伸縮手段14は、残差波形の伸縮を前記スペクトルパラメータに対応する区間について行うようにしてもよい。

【0091】また、実施の形態11は、実施の形態10におけるピッチ概形基準位置決定手段25を備えて、波形伸縮手段9がそのピッチ概形基準位置決定手段で得たピッチ概形基準位置の間隔が同じ話速変更率で伸縮されるように音声波形の伸縮率を設定するようにしてもよい。

【0092】また、実施の形態2、実施の形態3、実施の形態5～実施の形態8、実施の形態10、実施の形態11の波形伸縮手段9と、実施の形態4のVQ波形伸縮手段14と実施の形態9の母音中心部波形伸縮手段22は、母音および子音の伸縮比を規定する代りに、モーラ速度に対応する各母音および子音の伸縮比を記録するテーブルを参照するようにしてもよい。

【0093】また、実施の形態5と実施の形態6におけるリズム知覚点テーブル17、および実施の形態7におけるリズム知覚点兼伸縮比テーブル19は、音素の組み合わせのみならず、母音部のピッチの高さ、平均話速、文頭／文中／文末のどれに当たるかという情報に応じたリズム知覚点の値を記述することもできるようにしてもよい。

【0094】また、実施の形態5～実施の形態7におけるラベリング手段15は、マニュアルで音素ラベルを作成する代りに、簡易な音声認識手法によってラベリングを行うようにしてもよい。

【0095】

【発明の効果】以上のように、この発明によれば、学習音声データについてのコードベクトルを、各コードベクトルの母音および子音の出現確率とともに記録したVQコードブックを参照することにより、入力音声をベクトル量子化してコードベクトルの選択を行い、選択されたコードベクトルにおける母音または子音の出現確率によって判別された母音区間について、波形伸縮手段に与えられた話速変更率にて波形伸縮を行うように構成したので、母音区間についての波形伸縮を自動的に実施することが可能となり、全音声区間を同じ伸縮率で一様に変更することによる自然性劣化の問題を解消することができ、良好な発話速度変更を実現することができる話速変更装置が得られるという効果がある。

【0096】この発明によれば、母音区間と子音区間、または母音区間と子音区間と無音区間毎に固有の伸縮比を規定し、話速変更率が与えられると、規定された伸縮比に応じて求めた母音区間、子音区間、および無音区間の波形伸縮率に基づいて波形伸縮を行うように構成したので、母音区間と子音区間について個別の伸縮率で波形伸縮を自動的に実施することが可能となり、全音声区間を一様に同じ伸縮率で変更することによる自然性劣化の問題を解消することができ、良好な発話速度変更を実現することができるという効果がある。

【0097】この発明によれば、話速変更率が与えられた場合に、音素の種類や音素カテゴリー毎に規定された固有の伸縮比を記憶させた伸縮比テーブルを参照して求めた、入力音声の音素毎または音素カテゴリー毎の伸縮比に応じた区間毎の波形伸縮率に基づいて、波形伸縮を行うように構成したので、母音区間と子音カテゴリー毎に分類された区間について個別の伸縮率で波形伸縮を自動的に実施することができ、全音声区間を一様に同じ伸縮率で変更することによる自然性劣化の問題を解消することが可能となつて、良好な発話速度変更を実現することができるという効果がある。

【0098】この発明によれば、話速変更率が与えられた場合に、コードベクトルを各コードベクトルに固有の伸縮比率とともに記録しているVQコードブックの参照により、入力音声をベクトル量子化してコードベクトルの選択を行い、選択されたコードベクトル毎の伸縮比に応じて区間毎に求めた波形伸縮率に基づいて、波形伸縮を行うように構成したので、母音区間と子音区間についての個別の伸縮率での波形伸縮を簡便かつ自動的に実施することが可能となり、全音声区間を一様に同じ伸縮率で変更することによる自然性劣化の問題を解消することができ、良好な発話速度変更を実現することが可能に

なるという効果がある。

【0099】この発明によれば、音素系列毎に固有のリズム知覚点を記憶したリズム知覚点テーブルを参照して、入力音声よりラベリング手段が抽出した音素ラベルから入力音声のリズム知覚点を抽出し、話速変更率が与えられると、リズム知覚点間隔の伸縮率が一定になるように求めた区間毎の波形伸縮率に基づいて、波形伸縮を行うように構成したので、波形伸縮を自動的に実施することが可能となり、自然音声の時間構造の基本である、リズム知覚点を考慮した波形伸縮を行うことによって、同一伸縮率で一様に変換するよりも自然な発話速度変更を実現することができるという効果がある。

【0100】この発明によれば、母音区間と子音区間、または母音区間と子音区間と無音区間の各間毎に固有の伸縮比を規定し、話速変更率が与えられると、リズム知覚点間隔の伸縮率が一定になるように、規定された伸縮比に応じて求めた母音区間と子音区間、および母音区間と子音区間と無音区間の波形伸縮率に基づいて、波形伸縮を行うように構成したので、母音区間と子音区間について波形伸縮を自動的に実施することが可能となり、自然音声の時間構造の基本であるリズム知覚点を考慮した波形伸縮を行うことによって、一様に変換するよりも自然な発話速度変更を実現することができるという効果がある。

【0101】この発明によれば、話速変更率が与えられた場合に、音素の種類や音素カテゴリ毎に規定された固有の伸縮比を記憶している伸縮比テーブルを参照して、リズム知覚点間隔の伸縮率が一定になるように求めた、入力音声の音素または音素カテゴリ毎の伸縮比に応じた区間毎の波形伸縮率に基づいて、波形伸縮を行うように構成したので、母音区間と子音カテゴリ毎に分類された区間について、個別の伸縮率で波形伸縮を自動的に実施することが可能となり、自然音声の時間構造の基本であるリズム知覚点と音素種類を考慮した波形伸縮を行うことによって、一様変換よりも自然な発話速度変更を実現することができるという効果がある。

【0102】この発明によれば、音素系列に固有のリズム知覚点と音素系列に固有の伸縮比とを記憶するリズム知覚点兼伸縮比テーブルを参照して、入力音声よりラベリング手段が抽出した音素ラベルから入力音声のリズム知覚点と音素区間毎の伸縮比を抽出し、話速変更率が与えられると、リズム知覚点間隔の伸縮率が一定になるように求めた音素区間毎の波形伸縮率に基づいて、波形伸縮を行うように構成したので、母音区間と子音区間について、波形伸縮を自動的に実施することが可能となり、自然音声の時間構造の基本であるリズム知覚点と音素種類を考慮した波形伸縮を行うことによって、一様変換よりも自然な発話速度変更を実現することができるという効果がある。

【0103】この発明によれば、VQコードブックを参

照して抽出した母音区間音声の母音中心位置を決定し、話速変更率が与えられた場合に、その母音中心位置の間隔の伸縮率が一定になるように求めた、母音区間と子音区間、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率に基づいて波形伸縮を行うように構成したので、波形伸縮を自動的に実施することが可能となり、自然音声の母音間の間隔を利用できるように母音中心位置を考慮した波形伸縮を行うことによって、一様変換よりも自然な発話速度変更を実現することができるという効果がある。

【0104】この発明によれば、VQコードブックを参照して抽出した母音区間音声の母音中心位置を決定し、話速変更率が与えられた場合に、母音区間内の母音中心位置で最大または最小に、子音区間との境界に近付くとともに子音区間の波形伸縮率に近付くように決定した波形伸縮率に基づいて、各区間の波形伸縮を行うように構成したので、波形伸縮を自動的に実施することが可能となり、自然音声の母音間の間隔を利用できるように母音中心位置を考慮し、さらに子音との境界付近より母音中心位置周辺を強く伸縮するように重み付けした波形伸縮を行うことによって、一様変換よりも自然な発話速度変更を実現することができるという効果が得られる。

【0105】この発明によれば、入力音声のピッチ周波数分析で得られたピッチ楕形の極大値と極小値よりピッチ楕形基準位置間隔を決定し、話速変更率が与えられると、そのピッチ楕形基準位置間隔の伸縮率が一定になるように求めた、母音区間と子音区間、あるいは母音区間と子音区間と無音区間の伸縮比に応じた区間毎の波形伸縮率に基づいて、波形伸縮を行うように構成したので、波形伸縮を自動的に実施することが可能となり、自然音声のピッチ周波数より求めた基準位置に基づいて波形伸縮を行うことによって、ピッチ楕形を破壊することなく波形伸縮を行うことができ、一様変換よりも自然な発話速度変更を実現することが可能になるという効果がある。

【0106】この発明によれば、ピッチ周波数分析結果によるピッチ楕形の判定により、暫定子音区間音声を母音区間を含む部分と子音区間の部分とに分離して、改めて母音区間と子音区間との判別を行い、話速変更率が与えられた場合に、その母音区間と子音区間について固有に決定した伸縮比に応じて求めた、区間毎の波形伸縮率に基づいて波形伸縮を行うように構成したので、波形伸縮を自動的に実施することが可能となり、ピッチ楕形を母音子音区間判別の判定基準として利用することによって、ピッチ楕形を破壊せずに波形伸縮を行うことができ、一様変換よりも自然な発話速度変更を実現することが可能になるという効果がある。

【図面の簡単な説明】

【図1】 この発明の実施の形態1による話速変更装置の構成を示すブロック図である。

【図2】 この発明の実施の形態2による話速変更装置の構成を示すブロック図である。

【図3】 この発明の実施の形態3による話速変更装置の構成を示すブロック図である。

【図4】 この発明の実施の形態4による話速変更装置の構成を示すブロック図である。

【図5】 この発明の実施の形態5による話速変更装置の構成を示すブロック図である。

【図6】 この発明の実施の形態5におけるリズム知覚テーブルの内容を示す説明図である。

【図7】 この発明の実施の形態6による話速変更装置の構成を示すブロック図である。

【図8】 この発明の実施の形態7による話速変更装置の構成を示すブロック図である。

【図9】 この発明の実施の形態8による話速変更装置の構成を示すブロック図である。

【図10】 この発明の実施の形態9による話速変更装置の構成を示すブロック図である。

【図11】 この発明の実施の形態10による話速変更

装置の構成を示すブロック図である。

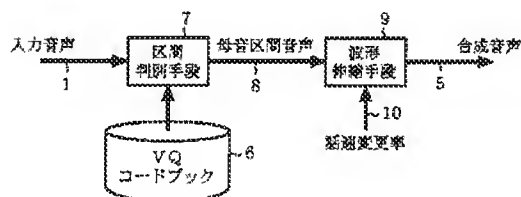
【図12】 この発明の実施の形態11による話速変更装置の構成を示すブロック図である。

【図13】 従来の話速変更装置の構成を示すブロック図である。

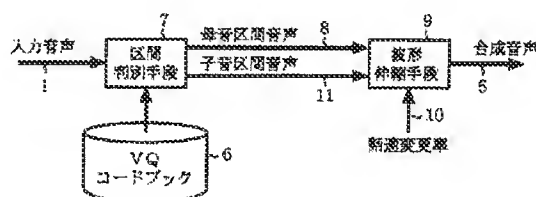
【符号の説明】

1 入力音声、5 合成音声、6 VQコードブック、7 区間判別手段、8 母音区間音声、9 波形伸縮手段、10 話速変更率、11 子音区間音声、12 伸縮比テーブル、13 伸縮比付きVQコードブック、14 VQ波形伸縮手段、15 ラベリング手段、16 音素ラベル、17 リズム知覚点テーブル、18 リズム知覚点抽出手段、19 リズム知覚点兼伸縮比テーブル、20 母音中心抽出手段、21 母音中心位置、22 母音中心部波形伸縮手段、23 ピッチ分析手段、24 ピッチ概形、25 ピッチ概形基準位置決定手段、26 ピッチ概形基準位置、27 暫定母音区間音声、28 暫定子音区間音声、29 ピッチ概形分割手段。

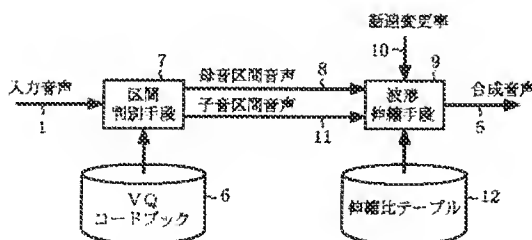
【図1】



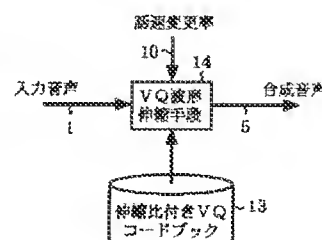
【図2】



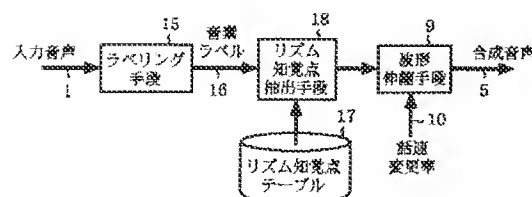
【図3】



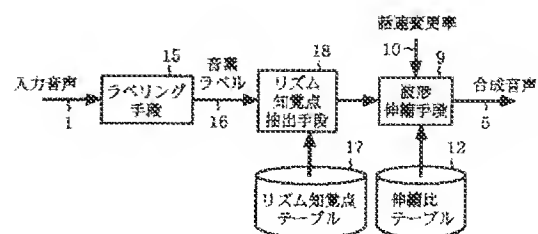
【図4】



【図5】



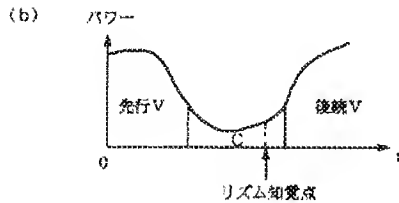
【図7】



【図6】

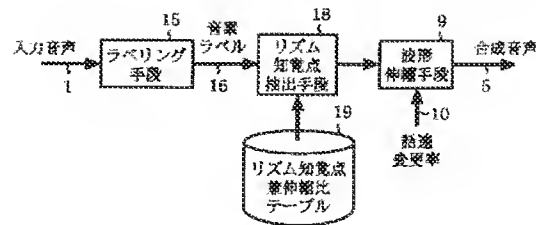
(a)

音素系列			リズム 知覚点
先行V	C	後続V	
...	k	a	7
a	k	a	10
i	k	a	10
u	k	a	10
e	k	a	10
o	k	a	10
N	k	a	10
...
...	r	a	4
a	r	a	7
i	r	a	7
u	r	a	6
...

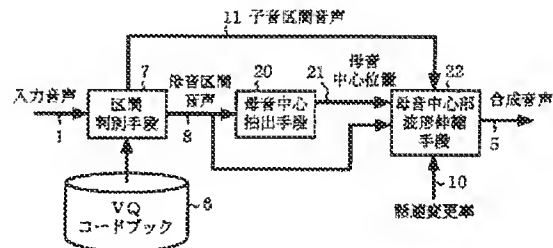


【図9】

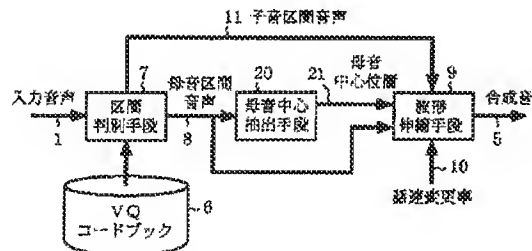
【図8】



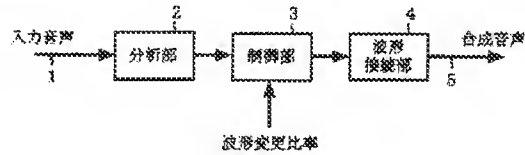
【図10】



【図13】



【図11】



【図12】

